Team Members

AKASH C0910253

SAKSHI C0908000

PRANIT C0913039

JAGADEESH C0912884

### **Table of Contents**

1. **Team Members** ...................................................................................................... *Page 1*
2. **Problem Statement** ................................................................................................ *Page 3*
3. **Approach to the Problem** ..................................................................................... *Page 3*
4. **Marketing Campaign Dataset Overview** .............................................................*Page 4*
   1. Dataset details (features and descriptions) ................................................. *Page 4*
5. **Imported Libraries and Their Purpose** ...............................................................*Page 5*
6. **Understanding the Dataset** ....................................................................................*Page 5*
   1. Basic dataset exploration commands .......................................................... *Page 5*
7. **Data Cleaning** ........................................................................................................ *Page 9*
   1. Checking for missing values........................................................................ *Page 9*
   2. Changing data types ...................................................................................*Page 10*
   3. Cleaning and converting key columns .......................................................*Page 10*
   4. Converting date formats .............................................................................*Page 11*
8. **Data Manipulation .............................................................................................** *Page 11* 
   1. Sorting data .............................................................................................. *Page 11*
   2. Filtering campaigns .................................................................................. *Page 11*
   3. Grouping data ............................................................................................ *Page 12*
9. **Performing the Exploratory Data Analysis (EDA)** .......................................... *Page 14*
10. **Correlation Analysis** ........................................................................................... *Page 15*
    1. Advertising Spend vs. Engagement Score ................................................ *Page 16*
    2. Advertising Spend vs. Conversion Rate ................................................... *Page 16*
    3. Advertising Spend vs. Acquisition Cost ................................................... *Page 16*
    4. Advertising Spend vs. ROI ....................................................................... *Page 17*
11. **Insights on Marketing Channel Performance** .................................................. *Page 18*
    1. Heatmap Analysis ...................................................................................... *Page 18*
    2. Key insights from the heatmap .................................................................. *Page 19*
    3. Best performing and underperforming marketing strategies ..................... *Page 19*
12. **References** ............................................................................................................. *Page 20*

**Problem statement**

Data Stream, a digital marketing agency, wants to improve customer insights by analyzing marketing campaign performance. The goal of this project is to build a simple script to import, manipulate, and analyze marketing data to find trends and correlations. The final output will be a basic report summarizing key findings.

### **Approach to the Problem**

1. Import the dataset from Kaggle.
2. Load and explore the dataset.
3. Data cleaning to eliminate missing values and format inconsistencies.
4. Conduct data manipulation such as sorting, filtering, and grouping.
5. Perform exploratory data analysis (EDA) with visualizations.
6. Compute correlations between data points.
7. Generate a summary report with key insights.

* We gathered the dataset from Kaggle *(Manisha Bhattacharjee, 2023)*.

### **Marketing Campaign Dataset Overview**

### This information is in line with the performance of different marketing campaigns. It holds data like target audience, time, and communication mediums. With 200,000 records and 16 features over two years, it helps analyze trends between different companies and consumer segments.

### **Dataset Details**

* **Company** – The brand or business running the campaign (fictional companies are included).
* **Campaign Type** – The kind of marketing approach used, such as email marketing, social media ads, influencer promotions, display ads, or search engine marketing.
* **Target Audience** – The specific group of people the campaign was aimed at, like women aged 25-34, men aged 18-24, or general audiences.
* **Duration** – The length of the campaign in days.
* **Channels Used** – The platforms where the campaign was promoted, such as email, social media, YouTube, websites, or Google Ads.
* **Conversion Rate** – The percentage of people who took the desired action after seeing the campaign, measuring its effectiveness.
* **Acquisition Cost** – The amount of money spent to acquire customers through the campaign.
* **ROI (Return on Investment)** – A measure of how profitable the campaign was based on the money earned compared to the amount spent.
* **Location** – The cities where the campaign ran, including major places like New York, Los Angeles, Chicago, Houston, and Miami.
* **Language** – The language used in the campaign, such as English, Spanish, French, German, or Mandarin.
* **Clicks** – The number of times people clicked on the campaign ads, showing engagement.
* **Impressions** – The total number of times the campaign was shown to people.
* **Engagement Score** – A rating from **1 to 10** that indicates how much interaction the campaign received.
* **Customer Segment** – The specific audience category the campaign targeted, such as tech lovers, fashion enthusiasts, health-conscious individuals, food lovers, or adventure seekers.
* **Date** – The date the campaign took place, allowing trend and pattern analysis over time.

### **Imported Libraries and Their Purpose**

* **pandas:** For handling and analyzing data *(Pandas, 2024).*
* **NumPy:** For numerical computations.
* **matplotlib:** For creating visualizations *(Matplotlib development team, 2025)*.
* **seaborn:** For advanced and attractive data visualizations.
* **warnings:** To ignore unnecessary warnings.

**Loading the Dataset:**

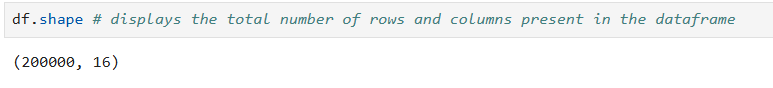
The dataset is loaded into a pandas Data Frame using read\_csv()



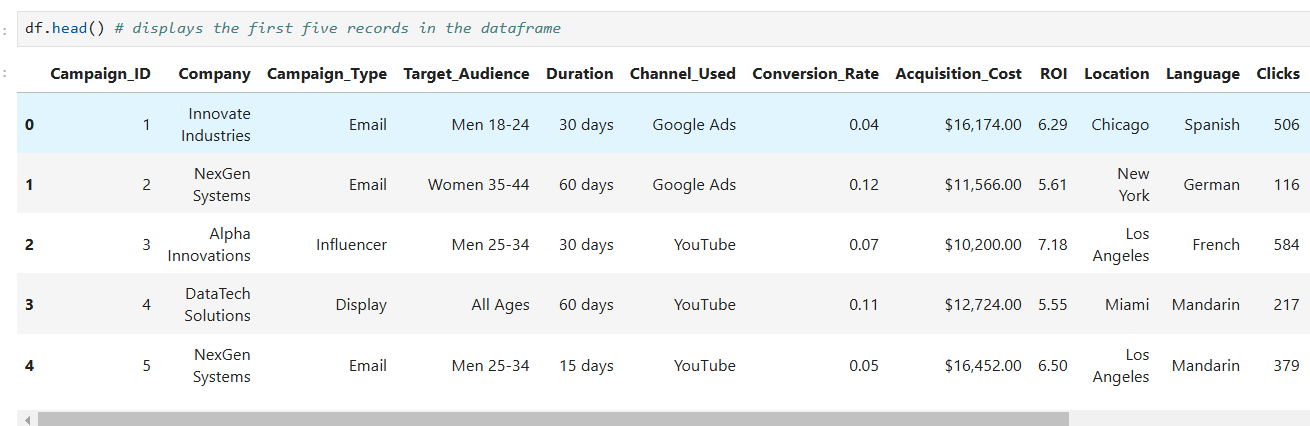
### **Understanding the Dataset**

### We need to understand what the dataset looks like before attempting to analyze it. We can do that with some simple commands:

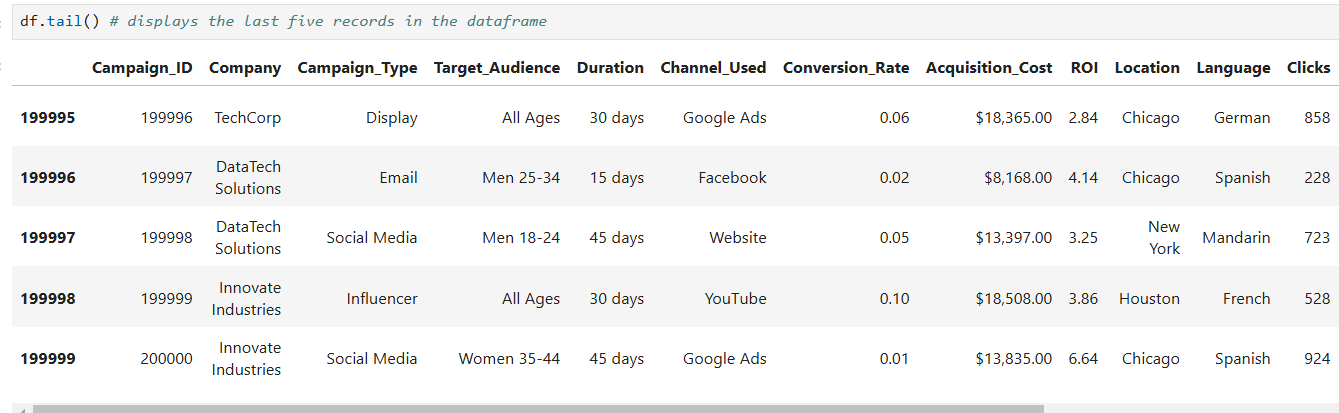
### **df.shape –** Tells us how many rows and columns the dataset has.



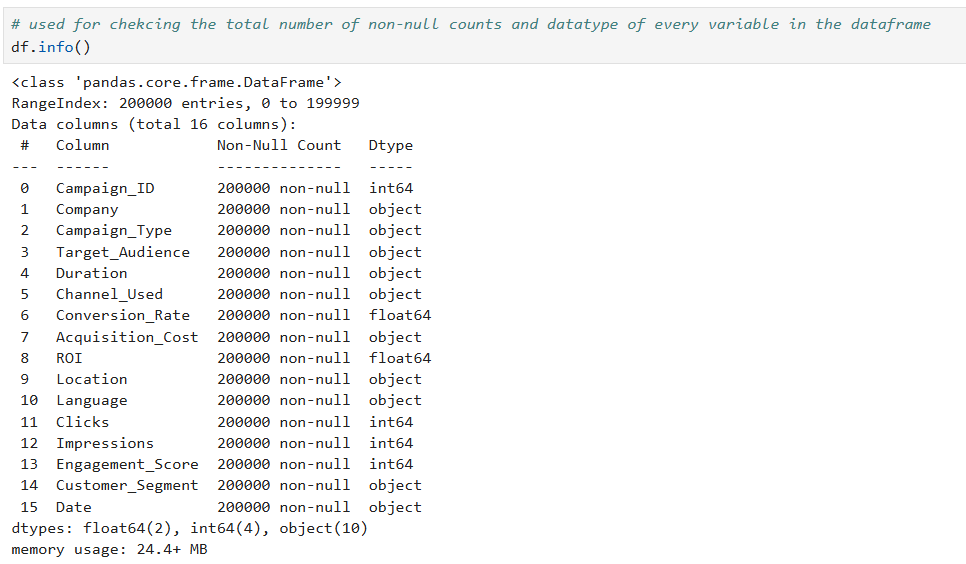
### **df.head() –** Displays the default last five rows so we have an idea about the data.



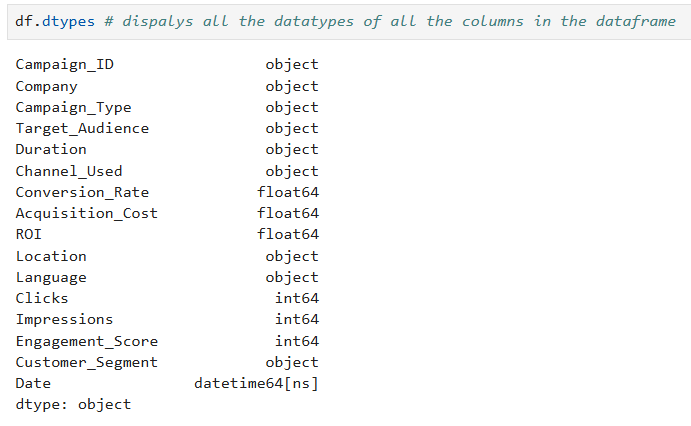
### **df.tail() –** Displays the default last five rows so that we know how the dataset ends.



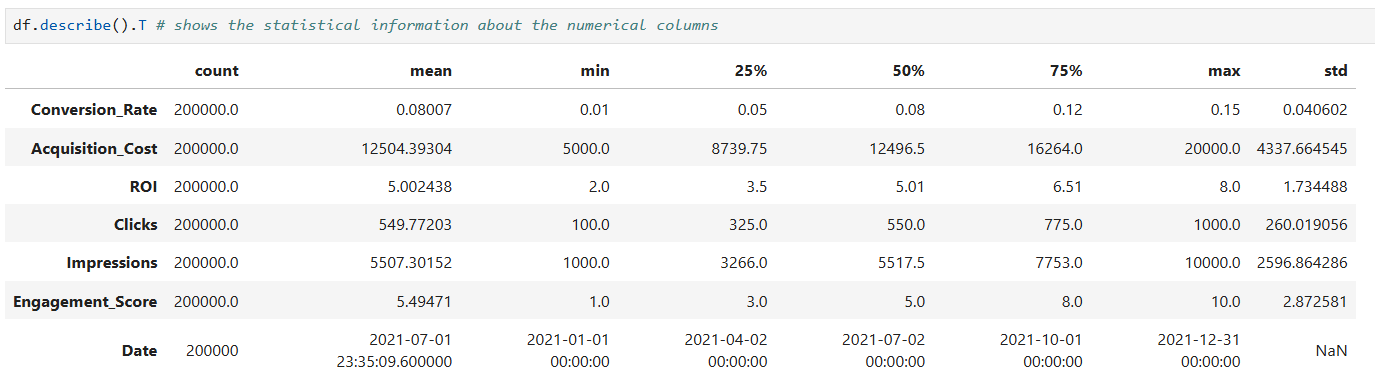
### **df.info () –** Provides info about all columns, including data types and missing values.



### **df.dtypes –** Shows the type of data in all columns, such as words or numbers.



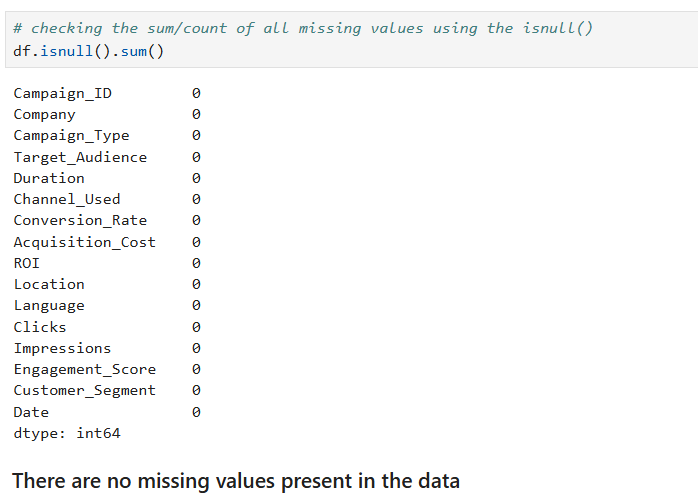
### **df.describe() –** Describes numerical data, showing key statistics like the average, the minimum, and the maximum.



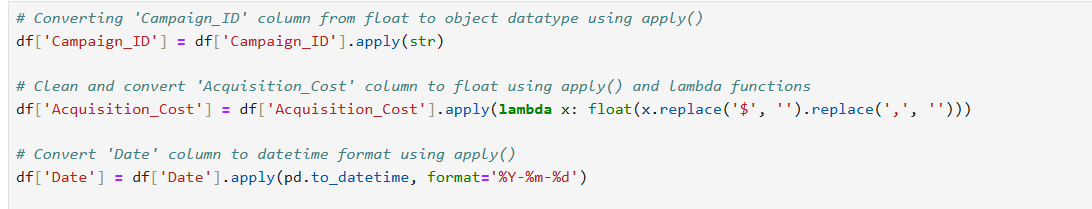
**Data Cleaning**

### Before analyzing the data, we need to check for missing values and ensure the data is in the correct format. Here’s what we did in this step:

1. **Checking for Missing Values**
   1. We used df.isnull(). sum () to check if any column has missing values.
   2. The output shows that **there are no missing values**, so we don’t need to fill in or remove any data.



1. **Changing the Data Type of 'Campaign\_ID'**
   1. The 'Campaign\_ID' column was originally in **float format**, which isn’t useful for an ID.
   2. We converted it to a **string (object data type)** using apply(str) because campaign IDs are identifiers, not numbers.
2. **Cleaning and Converting 'Acquisition\_Cost'**
   1. The 'Acquisition\_Cost' column contained dollar signs ($) and commas (,), making it a text value instead of a number.
   2. We used to apply (lambda x: float (x.replace('$', ''). replace (',', ''))) to remove symbols and convert it into a **numeric format (float)** so we can perform calculations on it.
3. **Converting 'Date' to Date Format**
   1. The 'Date' column was stored in string format earlier, but we required it to be in proper date format so that filtering, sorting, and trend analysis would be easier. We have performed this conversion using pd.to\_datetime()



**Data manipulation (Performing Sorting, Filtering, Grouping)**

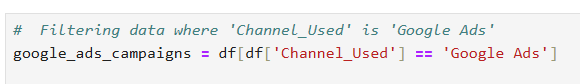
### **Sorting:**

### The data is sorted by 'Acquisition\_Cost' in descending order so we can easily see the most expensive campaigns.



### **Filtering:**

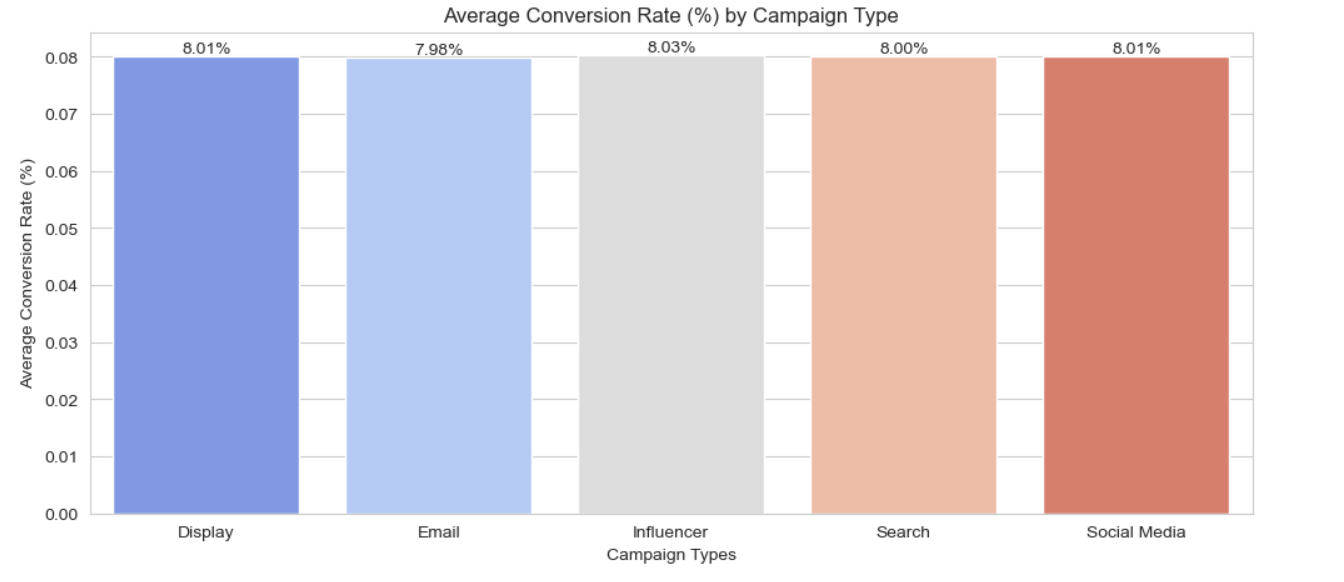
### We filtered the data to show only campaigns that used 'Google Ads' so we can look at those specific campaigns.



### **Grouping:**

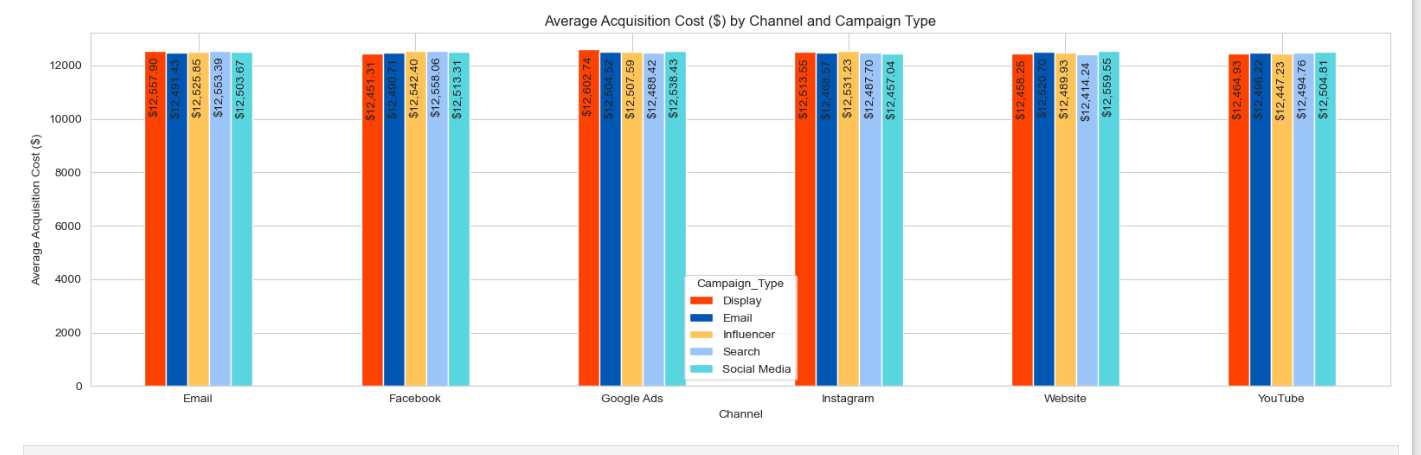
### The data is grouped by 'Campaign\_Type' so we can find the average conversion rate for each type so we can compare their performance.

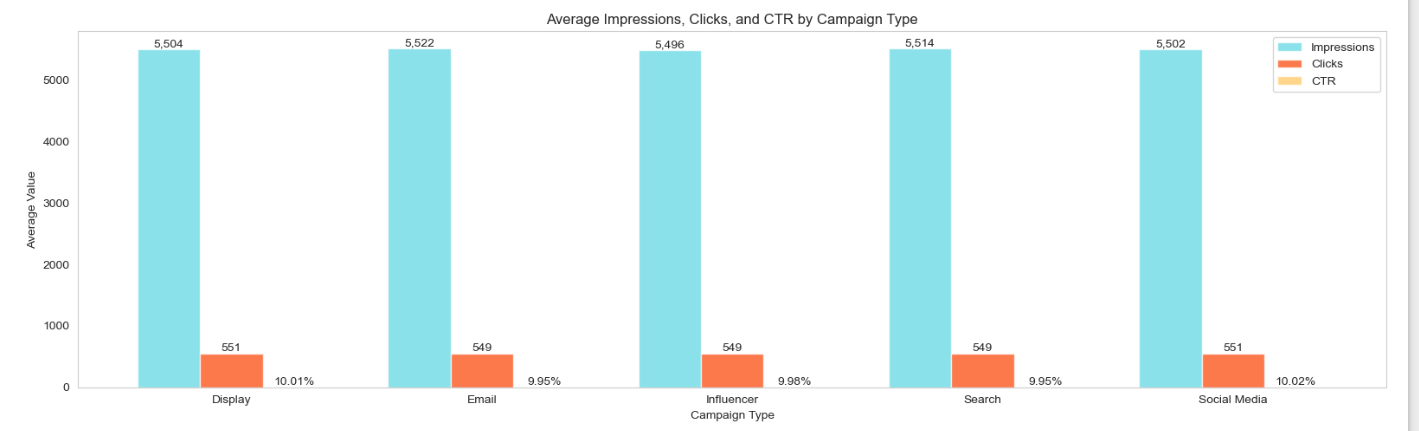


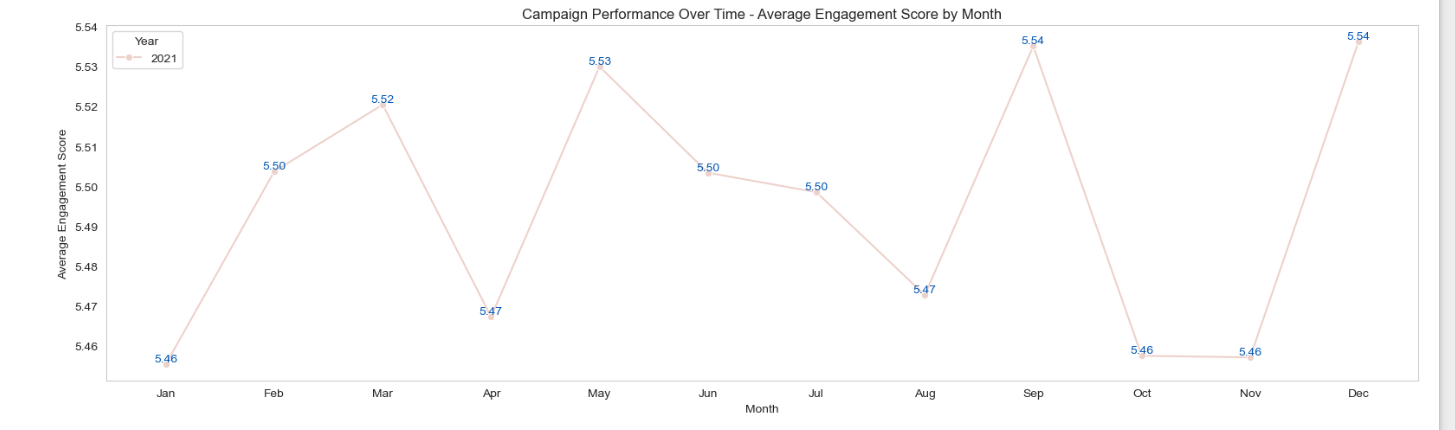


### The above bar chart depicts the average conversion rate for each type of campaign, and the values are displayed on the bars for simple comparison.

### **Performing the Exploratory Data Analysis**







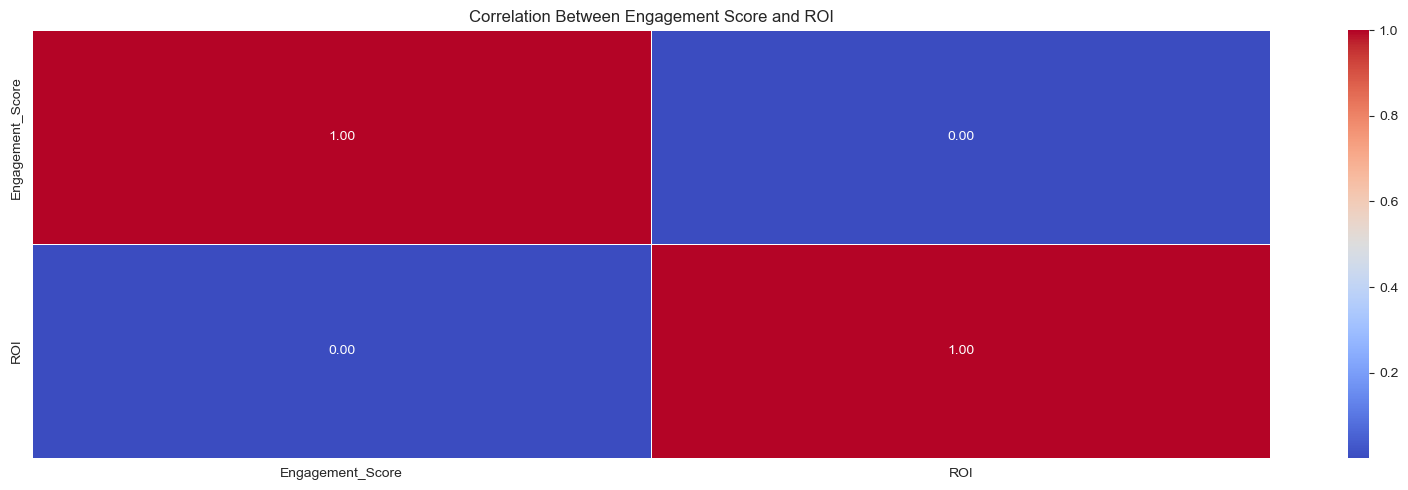
### **Correlation Analysis**

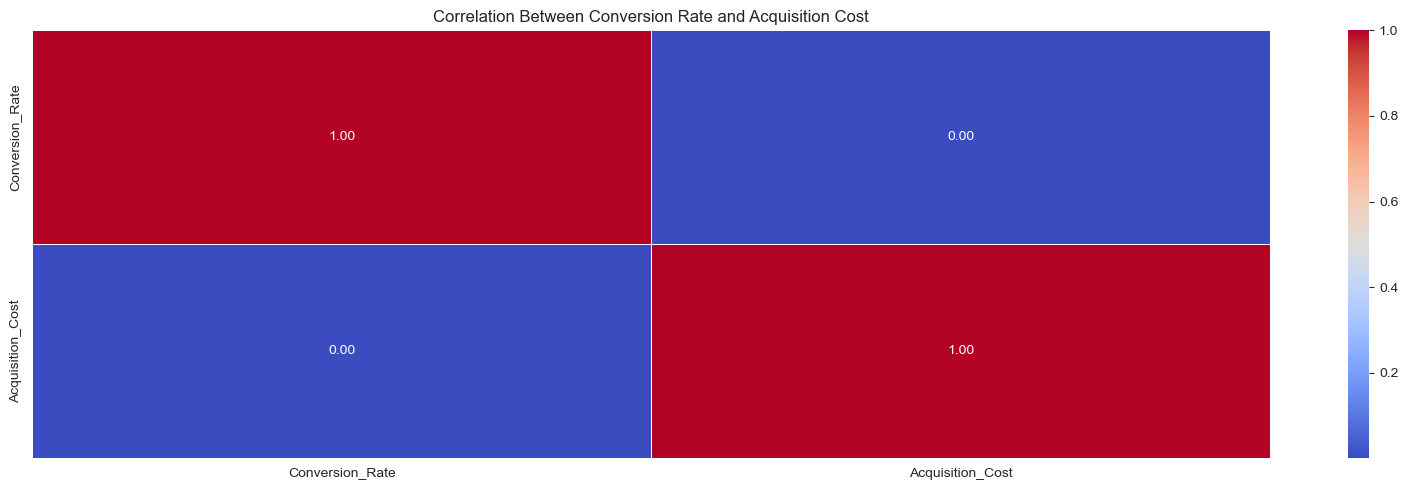
* To understand how advertising spending, customer engagement, and key performance metrics (like Conversion Rate, Acquisition Cost, and ROI) are related, we used basic correlation methods *(*[*Michael Waskom,*](https://mwaskom.github.io/) *2024)*. These methods help us see how strongly these factors are connected and in which direction.

Here’s what we did:

* **Correlation Coefficients:** We calculated the Pearson correlation to measure how much advertising spending is related to customer engagement and other key metrics.
* **Heatmap Visualization:** We created a heatmap to make it easier to see and compare the strength of these relationships between different metrics.







### **Key Findings**

### **Advertising Spend and Engagement Score:**

### We learned that there isn't a strong correlation between ad spend and customer engagement (Engagement Score). What this suggests is that an increase in ad spend does not directly correlate to increased levels of engagement.

### **Advertising Spend and Conversion Rate:**

### Similarly, advertising spend also does not display a high degree of correlation with Conversion Rate. This indicates that higher advertising spending does not lead to improved conversion rates.

### **Advertising Spending and Acquisition Cost:**

### Low correlation exists between advertising spend and Acquisition Cost such that higher spending on advertising will not necessarily yield lower customer acquisition costs.

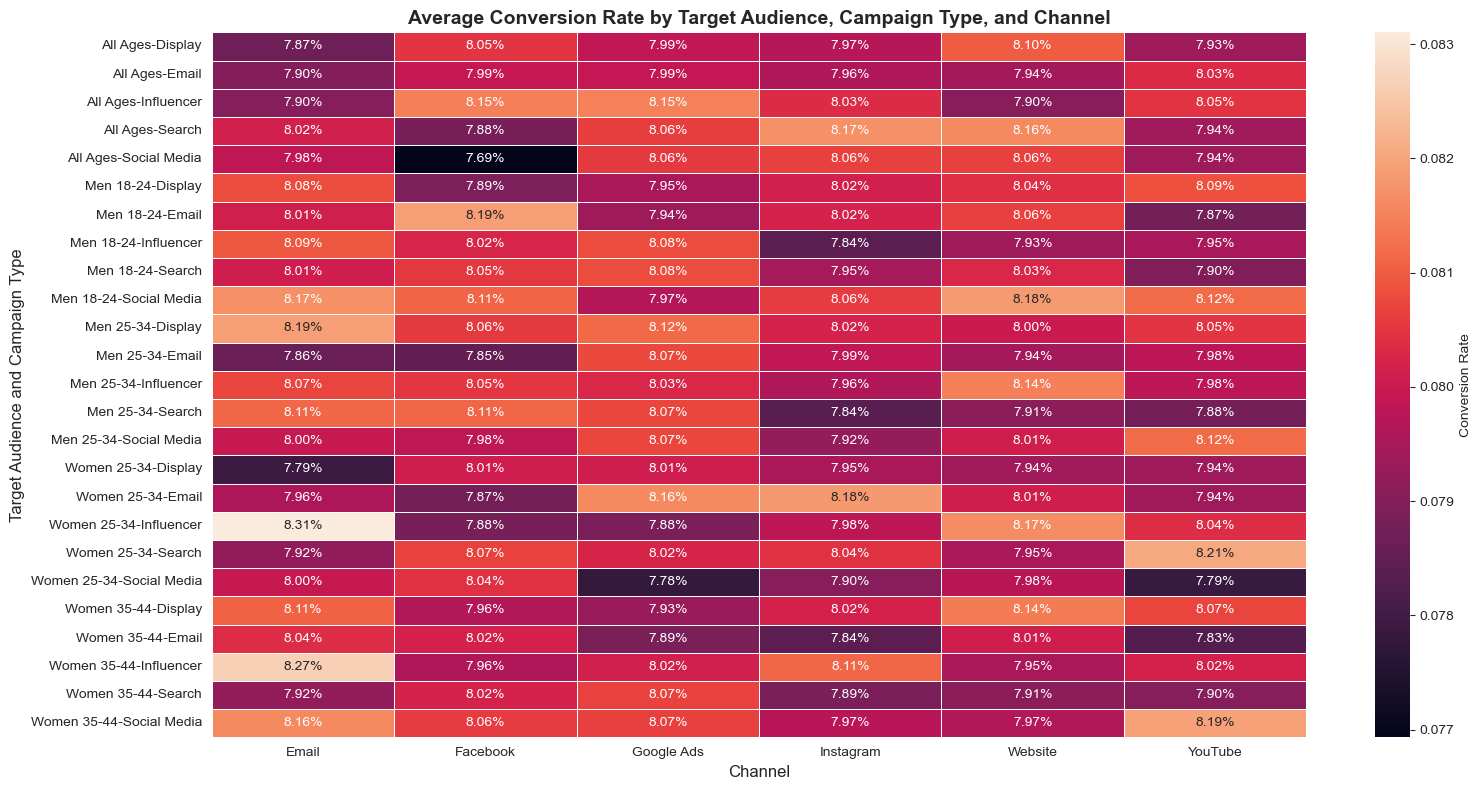
### **Advertising Spending and ROI:**

### There is no correlation between ad spend and ROI, so increasing or decreasing ad spend does not necessarily impact return on investment.

### Overall, our analysis indicates that advertising spend is not very correlated with customer engagement, conversion rates, acquisition cost, or return on investment. This would suggest that there are other drivers, outside of advertising expenditure, that have a significant impact on these metrics.

### **Insights on Marketing Channel Performance**

This analysis shows how different marketing channels perform across various customer groups. The goal is to identify which combinations of target audience, campaign type, and channel give the highest conversion rates, so marketers can focus on the most effective strategies.



#### **What the Heatmap Shows**

* **Y-Axis:** This shows different customer groups and the types of campaigns targeting them, like "Men 18-24 - Display Ads" or "Women 35-44 - Search Campaigns."
* **X-Axis:** This represents the marketing channels used in the campaigns, such as "Email," "Facebook," "Google Ads," "Instagram," "Website," and "YouTube."
* **Colors:** The colors in the heatmap tell us about the conversion rate. Lighter colors mean higher conversion rates, and darker colors mean lower conversion rates. The color scale on the side shows the exact conversion rate, ranging from about 7.7% to 8.3%.

#### **Key Insights**

1. **Best-Performing Combinations:**
   1. The highest conversion rate (8.19%) comes from the combination of **"Women 35-44 - Influencer on YouTube"**. This tells us that targeting this group with influencer marketing on YouTube is very effective.
   2. Another top performer is **"Men 18-24 - Social Media on Google Ads"**, which has a conversion rate of 8.18%.
   3. **"Women 35-44 - Search on Instagram"** also shows strong performance, with a conversion rate of 8.11%.
2. **Effective Campaign Channels:**
   1. Google Ads and Email are avenues with consistently high conversion rates, usually over 8%, across various customer segments and campaign types. These are good marketing avenues to target.
3. **Underperforming Combinations:**
   1. The lowest conversion rates are found in darker sections of the heatmap, particularly in channels like **social media on Facebook and Instagram**, targeting the **"All Ages"** group. These combinations show conversion rates between 7.69% and 7.98%, which are lower than other channels.

### Overall, Thís heatmap helps marketers understand which combinations of target audience, campaign type, and channel give the best results. By focusing on the most successful strategies (like YouTube influencer marketing for women aged 35-44), businesses can improve their marketing efforts and achieve higher conversion rates. At the same time, knowing which combinations aren't working as well (like Facebook or Instagram for the "All Ages" audience) can help marketers avoid wasting resources on underperforming tactics.

References:

* Manisha Bhattacharjee, 2023,

<https://www.kaggle.com/datasets/manishabhatt22/marketing-campaign-performance-dataset/data>

* Pandas, 2024,

<https://pandas.pydata.org/docs/>

* Matplotlib development team, 2025,

<https://matplotlib.org/stable/index.html>

* [Michael Waskom,](https://mwaskom.github.io/) 2024,

<https://seaborn.pydata.org/>